PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2012) **279**, 1943–1949
doi:10.1098/rspb.2011.2181
Published online 4 January 2012

How do communication systems emerge?

Thomas C. Scott-Phillips[1,*], Richard A. Blythe[2], Andy Gardner[3,4] and Stuart A. West[3]

[1]*School of Psychology, Philosophy and Language Sciences, University of Edinburgh,*
3 Charles Street, Edinburgh EH8 9AD, UK
[2]*School of Physics and Astronomy, University of Edinburgh, James Clerk Maxwell Building,*
Mayfield Road, Edinburgh EH9 3JZ, UK
[3]*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK*
[4]*Balliol College, University of Oxford, Broad Street, Oxford OX1 3BJ, UK*

Communication involves a pair of behaviours—a signal and a response—that are functionally interdependent. Consequently, the emergence of communication involves a chicken-and-egg problem: if signals and responses are dependent on one another, then how does such a relationship emerge in the first place? The empirical literature suggests two solutions to this problem: ritualization and sensory manipulation; and instances of ritualization appear to be more common. However, it is not clear from a theoretical perspective why this should be the case, nor if there are any other routes to communication. Here, we develop an analytical model to examine how communication can emerge. We show that: (i) a state of non-interaction is evolutionarily stable, and so communication will not necessarily emerge even when it is in both parties' interest; (ii) the conditions for sensory manipulation are more stringent than for ritualization, and hence ritualization is likely to be more common; and (iii) communication can arise by a third route, when the intention to communicate can itself be communicated, but this may be limited to humans. More generally, our results demonstrate the utility of a functional approach to communication.

Keywords: communication; signals; evolution; emergence; communicative intention

1. INTRODUCTION

Communication is not a trait possessed by one or another individual. Rather, it is an interaction between two (or more) individuals [1–4]. This is reflected in contemporary definitions of communication, which emphasize that whether a given behaviour is a signal depends on whether there is a corresponding response, and vice versa [3,4]. This suggests a chicken-and-egg problem: if signals and responses depend on each other to explain their adaptive value, then how can communication emerge in the first place?

The existing literature suggests two broad processes by which communication can arise: ritualization and sensory manipulation [3,5]. In ritualization, signals evolve from behaviours that were originally only cues. For example, the use of urine to mark territory may have begun as a marker of fear, produced by animals at the periphery of territory in which they felt safe, which other animals used as the cue of the focal individual's presence [3]. In sensory manipulation, signals evolve from behaviours that were originally only coercive. For example, many mating displays may have begun as scenarios in which a preference for objects of a certain colour allowed the behaviour of potential receivers to be manipulated by others [5]. In the empirical literature, examples of ritualization are more common than accounts of sensory manipulation, and it has been suggested that 'most signals probably evolved by . . . ritualization' ([3], p. 68).

These empirical observations raise (at least) three questions. First, why is ritualization more common? Is there a particular theoretical point that might explain this? Second, are ritualization and sensory manipulation the only ways in which communication can emerge? Are there any other routes to communication? Third, how do these two routes to communication (and any others that might be identified) relate to one another, and to the way in which communication is defined? In other words, are there general patterns in the ways in which communication systems emerge from non-communicative states, and can these be predicted from how we define communication?

In this paper, we examine theoretically the different ways in which communication systems can emerge from states of no communication. In particular, we explore the different ways in which the chicken-and-egg interdependence of signals and responses can emerge, and we ask about the relative frequency of the possible routes we describe. Although our work is inspired by the evolution of animal signals, our larger goal is to develop a general framework that can be applied more broadly. For example, our results also hold for the ontogenetic emergence of communication, such as that between pairs of interacting primates [6,7]. It is for this reason that we begin by specifying exactly what we mean by communication, and associated terms and concepts (§2). We then describe our model and main results (§§3–5). The model is deliberately simple, since it is designed to illustrate general functional principles about how communication systems emerge, rather than the mechanisms involved in any particular instance.

2. A DEFINITION OF COMMUNICATION

Since our objective is to study the emergence of communication from a state of non-interaction, it is necessary that

[*] Author for correspondence (thom@ling.ed.ac.uk).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2011.2181 or via http://rspb.royalsocietypublishing.org.

Received 17 October 2011
Accepted 6 December 2011

Table 1. Definitions of signals, responses, cues and coercion. This table makes clear the relationship that cues and coercion have with communication: communication can be thought of as an interaction that is both a cue and a coercive behaviour. Note also that these are general definitions, defined in terms of functionality, and as such are applicable to any instance of communication, and not only animal signals.

	function of action to affect receiver?	function of reaction to be affected by the action?
communication	yes	yes
cue	no	yes
coercion	yes	no

we are able to identify scenarios that are communicative, and distinguish them from scenarios that are not. We define a signal as any action or structure that causes a reaction in another organism, where it is the function of both action and reaction to play these particular roles in the interaction [3,4]. If these conditions are satisfied, then the action is a signal; the reaction is a response; and the overall interaction is communicative. If only the reaction is functional in this way, then the action is a cue; and if only the action is functional in this way, then it is coercive (table 1). (Note: the term *coercive* does not imply that the interaction is not beneficial for the reacting organism. It may indeed be beneficial. All that *coercion* implies is that the reaction did not evolve as part of the interaction.) These definitions capture various *prima facie* instances of communication, and appropriately exclude phenomena that we would not wish to term communicative, such as camouflage [3,4]. Furthermore, these definitions make clear that signals and their corresponding responses are interdependent: both are required for an interaction to be communicative. In the electronic supplementary material, we discuss how these concerns relate to the role of cooperation in communication, and to other issues in animal signalling theory, in particular the matter of honesty.

3. ANALYTICAL MODEL

(a) *Basic set-up*

Our basic model involves two individuals: an actor and a reactor. At this stage, we do not label the individuals as signaller and receiver, because we want to investigate the conditions under which behaviours do and do not become signals and responses. At the point at which action and reaction satisfy the functional criteria of our definition above, we will label them as signal and response accordingly.

The world can be in one of n possible states, $T = \{t_1, t_2, \ldots, t_n\}$. The state of the world is known to the actor but not to the reactor. Each state t_i occurs with a fixed, positive probability, $\varphi(t_i) > 0$ (so $\sum_i \varphi(t_i) = 1$). Whatever the state of the world, the actor can perform one action from a set, $A = \{a_0, a_1, \ldots\}$, and the reactor can perform one reaction from a different set, $R = \{r_0, r_1, \ldots\}$. Note that the sets of possible actions and possible reactions include a_0 and r_0, respectively: these refer to the actor/reactor doing nothing, relative to whatever behaviour they were already engaged in. Consider, for example, prey fleeing from a predator. Here, the fleeing is the default, non-signalling behaviour (a_0), and in the absence of communicative considerations this will be optimized according to factors such as the expected length of pursuit, the need to conserve energy, and so on. However, animals may run faster than this optimal speed, in order to advertise an ability to escape, and hence deter the predator from continuing [8]. It is these possible deviations from a_0 that comprise the remainder of A, the set of possible actions available to the actor. R, the set of possible reaction, is characterized in the same way.

Then for each pair of states of the world and reactions, there will be a pair of payoffs, one each for actor and reactor: $\Pi_{\mathrm{A}}(t_i, r_k)$ and $\Pi_{\mathrm{R}}(t_i, r_k)$. These payoffs are measured relative to the scenario in which there is no interaction between actor and reactor. Consequently the payoffs associated with the reactor doing nothing are fixed at 0: $\Pi_{\mathrm{A}}(t_i, r_0) = \Pi_{\mathrm{R}}(t_i, r_0) = 0$. In addition, there is an efficacy cost associated with all behaviours except those that involve doing nothing, to reflect the energy expenditure in performing the behaviour in question. This cost can be different for different actions and reactions (i.e. $\forall\ a_j$ where $j \neq 0$, $\exists$ cost $\varepsilon(a_j) > 0$, $\varepsilon(a_0) = 0$; and $\forall\ r_k$ where $k \neq 0$, $\exists$ cost $\varepsilon(r_k) > 0$, $\varepsilon(r_0) = 0$).

Following Donaldson *et al.* [9], we then define the actor's strategy as a matrix of the conditional probabilities that the actor will perform a particular action, given each particular state of the world ($\mathbf{P} = p(a_j \mid t_i)$; $\sum_j p(a_j \mid t_i) = 1\ \forall\, i$). Similarly, we define the reactor's strategy as a matrix of the conditional probabilities that the reactor will perform a particular reaction, given each particular action ($\mathbf{Q} = q(r_k \mid a_j)$; $\sum_k q(r_k \mid a_j) = 1\ \forall\, j$).

To establish the net payoff to the actor, we first calculate the product of: (i) the probability that the world is in a particular state t_i; (ii) the probability that the actor will perform a particular action, a_j, given that the world is in state t_i; (iii) the probability that the reactor will perform a particular reaction, r_k, given that the actor has performed a_j; and (iv) the payoff for the actor associated with the particular combination of state and reaction. We then sum this product over all possible states, actions and reactions, and subtract any efficacy cost associated with the performance of the action. This gives us

$$w_{\mathrm{A}}(P, Q) = \sum_{t\in T}\sum_{a\in A}\sum_{r\in R} \varphi(t)p(a|t)q(r|a)\Pi_{\mathrm{A}}(t|r) - \sum_{a\in A} \varepsilon(a)p(a),$$

where $p(a)$ is the weighted sum over T of the conditional probabilities $p(a|t)$. Similarly, the net payoff to the reactor is

$$w_{\mathrm{R}}(P, Q) = \sum_{t\in T}\sum_{a\in A}\sum_{r\in R} \varphi(t)p(a|t)q(r|a)\Pi_{\mathrm{R}}(t|r) - \sum_{r\in R} \varepsilon(r)q(r),$$

where $q(r)$ is the weighted sum over A of the conditional probabilities $q(r|a)$.

Since we wish to understand how communication can emerge from a state of no communication, we must consider strategies that correspond to no interaction between actor and reactor. To do this, we define a null matrix for each, which corresponds to doing nothing regardless of what the other individual does. So $\mathbf{P}_{\mathrm{null}}$ is defined by the actor performing a_0 with probability 1 for all states of the world (so $p(a_0|t_i) = 1 \,\forall\, I$, and hence $p(a_j|t_i) = 0 \,\forall\, j \neq 0$). Similarly $\mathbf{Q}_{\text{null}}$ is defined by the reactor performing r_0 with probability 1 for all actions (so $q(r_0|a_j) = 1 \,\forall\, j$, and hence $q(r_k|a_j) = 0 \,\forall\, k \neq 0$).

Finally, we also assume that if the actor does nothing, then the reactor's best strategy is to do nothing as well, such that if the actor provides no information about the state of the world, then the reactor's best strategy is to do nothing, rather than to perform a behaviour at random ($w_{\text{R}}(\mathbf{P}_{\text{null}}, \mathbf{Q}_{\text{null}}) \geq w_{\text{R}}(\mathbf{P}_{\text{null}}, \mathbf{Q}') \,\forall\, \mathbf{Q}' \neq \mathbf{Q}_{\text{null}}$).

(b) *A state of non-interaction is evolutionarily stable*

We can now ask whether (and if so, how) communication might evolve from a wholly non-interactive initial scenario in which there is no interaction (i.e. from the state ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$)). It should be relatively obvious that, starting from a state of no interaction (i.e. where the actor always chooses a_0 and the reactor always chooses r_0), any unilateral change in strategy will not increase either individual's payoff. If the actor unilaterally changes strategy from always doing nothing then the only difference to their payoff will be the efficacy cost that is associated with all actions except for a_0; there will be no additional benefit because the reactor will always ignore them. Correspondingly, if the reactor unilaterally changes strategy from always ignoring the actor (i.e. from always choosing r_0), then their payoff will necessarily be less than zero, since it is an assumption of the model that if the actor does nothing, then the reactor's best strategy is to do nothing as well. In other words, the pair of strategies ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$) is evolutionarily stable. This result is intuitive, and we prove it formally in the electronic supplementary material. Moreover, it has long been known, especially from research on begging, that many communication games have stable non-signalling equilibria [10,11]. However, the implications of this for general patterns of how communication systems emerge have not previously been spelt out explicitly.

The immediate corollary is that we cannot simply assume that if communication is beneficial for both parties it will necessarily emerge. Our model shows why such an assumption is naive: communication is an inherently interdependent phenomenon, and this interdependence imposes constraints on the dynamics by which communication can emerge. Specifically: both signals and responses depend on each other for their adaptive value, and this makes the emergence of communication a chicken-and-egg problem. The next section considers how this problem can be overcome.

4. RITUALIZATION AND SENSORY MANIPULATION

The empirical literature suggests that communication evolves by one of two processes: ritualization and sensory manipulation [3,5]. As mentioned in §1, a possible example of ritualization is the use of urine and faeces to mark territory [12]. Here, the territory owner initially relieves himself because of fear, but at the same time he is willing to remain and fight for the territory. Hence, the urine and faeces act as cues to others about the ownership of the territory, which may change their behaviour accordingly. The focal individual may then evolve to urinate (or defecate) in order that others recognize ownership of the territory, whether or not he is scared. So here a cue evolved first, and was then co-opted by the (proto-)signaller, and hence became a signal. An example of sensory manipulation may be the offering of nuptial gifts, from males to females, that occurs in many insect species (see Vahed [13] for a review). A specific example is the scorpionfly *Bittacus apicalis*, where males capture large prey and then offer it to females who feed on it during copulation [14]. The offering of prey is a signal, which may have initially involved the male simply presenting the food to the female, who is willing to mate because she has a pre-existing mechanism that prioritizes the opportunity to feed on large prey. At this point, the presentation of the prey is coercive. If there is later positive selection on the female to accept the prey in exchange for copulation, then it has become a signal [3]. These two processes are summarized in figures 1 and 2, respectively. We now formally model each, as a way to specify the similarities and differences between them. This will allow us to ask about their relative frequency (§4*c*), and also whether there are any other ways in which communication can emerge (§5).

(a) *Ritualization*

We first examine how signals can evolve via ritualization, in which signals evolve from preceding cues (figure 1). To do this, we must first specify how the initial conditions differ from a state of total non-interaction (since, as shown in §3*b* above, communication is unlikely to emerge from such a state). We hence state that one particular state of the world, t_I, has the following properties. First, if a particular action, a_J, is performed when the world is in this state, then there is a positive payoff, α, for the actor, independent of any effect that action may have as a result of its effect upon the reactor. This is equivalent to the production of urine owing to fear in the example discussed above. Second, we also specify that in the same state of world, there is a particular reaction, r_K, that produces a positive payoff for the reactor ($\Pi_{\text{R}}(t_I, r_K) > 0$). This is equivalent to rival individuals being able to use the presence of urine as a guide to the ownership of the territory, and hence behave accordingly.

What are the selection consequences of these changes? The actor will evolve to perform a_J whenever the world is in state t_I. The reactor will then, in turn, evolve to perform r_K whenever the actor performs a_J. Thus, at this point, a_J is a cue: it has an effect upon the reactor, the reactor has evolved a reaction to attend to it, but the actor has not evolved to cause that reaction.

The evolution of the cue may, in turn, have evolutionary consequences for the actor. These may be negative, neutral or positive (i.e. $\Pi_{\text{A}}(t_I, r_K)$ may be less than, equal to or greater than, 0). If the consequences are negative, and if they outweigh the benefit that the actor receives for performing the behaviour in this state of the world, then there will be selection for the actor not to perform that action any more (i.e. if $\Pi_{\text{A}}(t_I, r_K) < 0$ and if $-\Pi_{\text{A}}(t_I, r_K) < \alpha$). If the consequences are neutral, or if they are negative but are outweighed by the benefit that the actor receives for performing the behaviour in this state of the world, then there will be no selection on the actor (i.e. if $\Pi_{\text{A}}(t_I, r_K) = 0$ or if $\Pi_{\text{A}}(t_I, r_K) < 0$ but

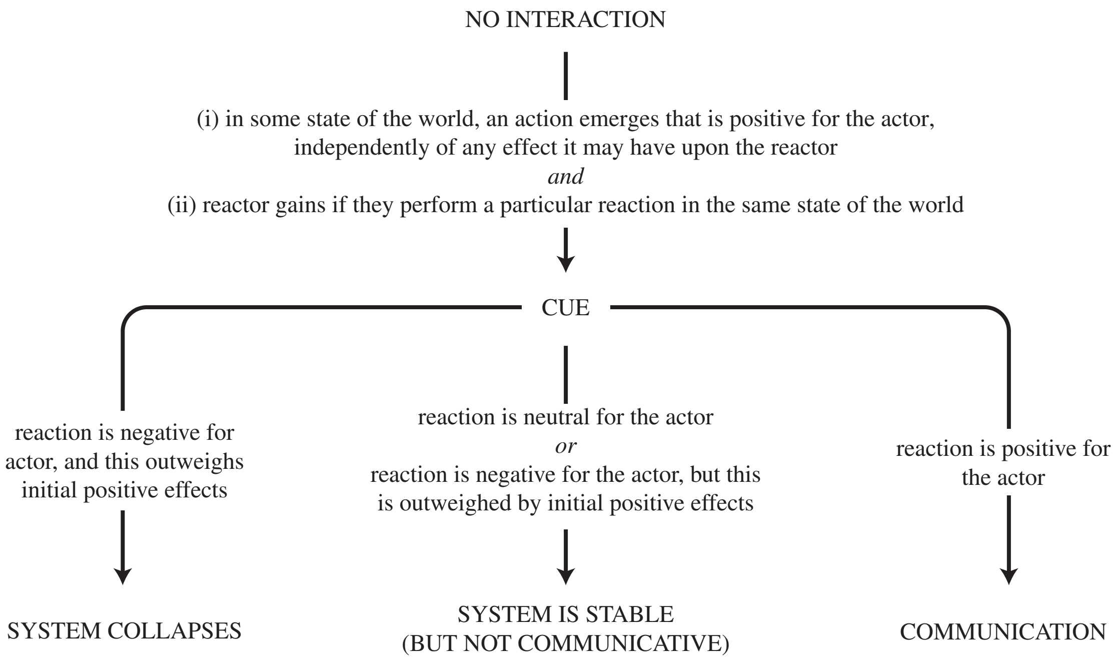
NO INTERACTION
(i) in some state of the world, an action emerges that is positive for the actor, independently of any effect it may have upon the reactor
and
(ii) reactor gains if they perform a particular reaction in the same state of the world
CUE
reaction is negative for actor, and this outweighs initial positive effects
reaction is neutral for the actor
or
reaction is negative for the actor, but this is outweighed by initial positive effects
reaction is positive for the actor
SYSTEM COLLAPSES
SYSTEM IS STABLE (BUT NOT COMMUNICATIVE)
COMMUNICATION

Figure 1. Ritualization. Ritualization involves two stages. First, a cue emerges. Then, that cue may become a signal, and the interaction may become communicative, if the cueing individual (the actor) gains from their production of the cue. In contrast, if the production of the cue is costly for the actor, then the system will collapse. If it is neutral, then the system will remain stable, but we cannot label it communicative.

$-\Pi_A(t_I, r_K) > \alpha$). If the consequences are positive, then the actor's behaviour will be maintained under positive selection (i.e. if $\Pi_A(t_I, r_K) > 0$).

Only the last of these scenarios is communicative: it is only here that the action's effects upon the reactor explain (in part) its continued existence (recall from §2*a* that this is a necessary criterion for something to be the function of behaviour). We can now term the action a signal, and the reaction a response. Moreover, this state has emerged via a process of ritualization: a cue has become a signal. In the other two scenarios, the action has either been selected against, or it has been maintained, but not because of its effects upon the reactor. Hence, neither scenario is communicative.

(b) *Sensory manipulation*

We now examine how signals can evolve via sensory manipulation, in which signals evolve from preceding coercive behaviours (figure 2). As with ritualization, we must first specify how the initial conditions differ from a state of non-interaction. We first specify that there is a particular action, a_J, that (because of some pre-existing mechanism) produces the reaction r_K ($q(r_K|a_J) = 1$). Translated into the scorpionfly example, the action is the presentation of prey by the male, and the reaction is the female feeding on it. Second, we specify that for one particular state of the world, t_I, there is a positive payoff to the actor if the reactor performs r_K ($\Pi_A(t_I, r_K) > 0$). Again translated into the scorpionfly example, this is equivalent to the male being able to mate if the female is feeding.

What are the selection consequences of these changes? First, the actor will evolve to perform a_J whenever the world is in state t_I, since that will produce reaction r_K, which has a positive payoff for the actor. Thus at this point, a_K is coercive: it has an effect upon the reactor, this effect is the function of the action, but the reaction is not functional.

This development will, in turn, have selection consequences for the reactor. These may be negative, neutral or positive (that is, $\Pi_R(t_I, r_K)$ may be less than, equal to or greater than 0). If the net consequences are negative, then there will be selection for the reactor not to perform that action any more (i.e. if $\Pi_R(t_I, r_K) < 0$), and the system will collapse. If the consequences are neutral, then there will be no selection on the reactor (i.e. if $\Pi_R(t_I, r_K) = 0$). An example that illustrates the difference between these would be mimicry, in which the actor mimics, say, a female in order to attract prey. This is an act of coercion, and it is costly for the reactor. If it is sufficiently common to outweigh the benefits of being attracted to females, then the net consequences are negative, and the prey will evolve a defence mechanism of some sort. If, on the other hand, these costs are balanced by the mating opportunities that follow from being attracted to females, then the net consequences are neutral.

If the net consequences are positive, then the reactor's behaviour will be maintained under positive selection (i.e. if $\Pi_R(t_I, r_K) > 0$). As before, it is only this final scenario that is communicative: it is only here that the action's effects upon the reactor explain (in part) its continued existence, and so it is only here that we can term the reaction a response. This state has emerged via a process of sensory manipulation: a previously coercive behaviour has become a signal. In the other two scenarios, the action has either been selected against, and hence the interaction collapses; or the action has been maintained, but not because of its effects upon the reactor. Hence, neither is communicative.

It is clear from these models that ritualization and sensory manipulation are closely related. However, our

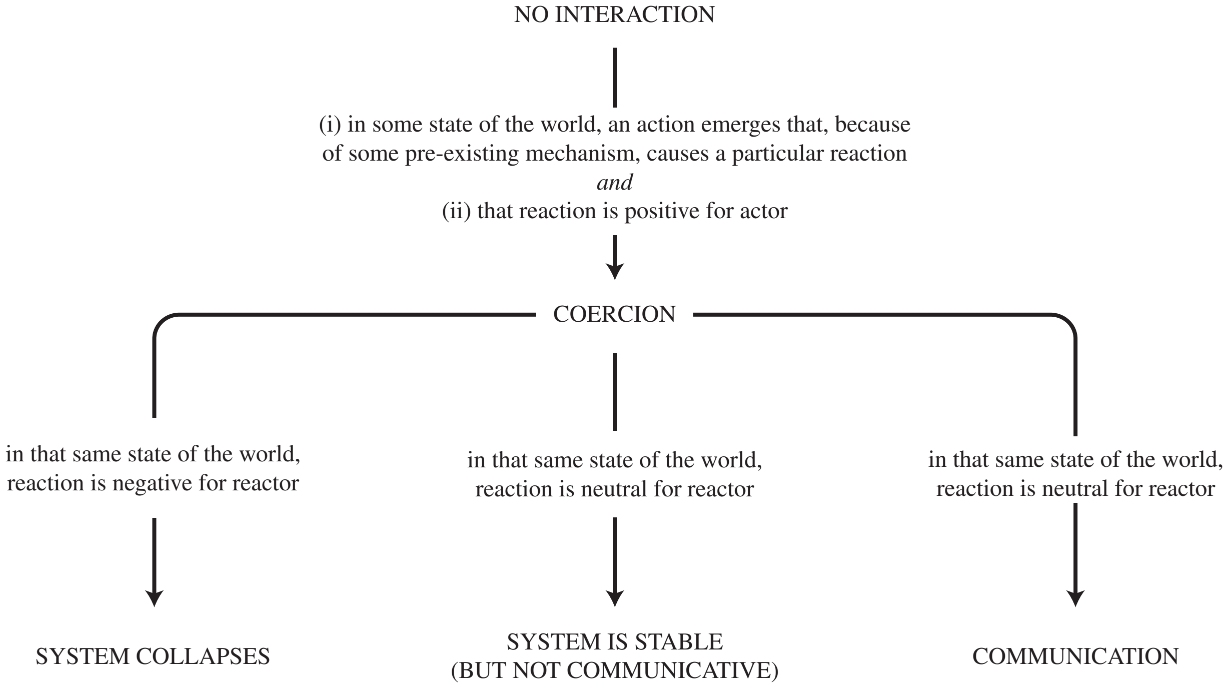
NO INTERACTION
(i) in some state of the world, an action emerges that, because of some pre-existing mechanism, causes a particular reaction
and
(ii) that reaction is positive for actor
COERCION
in that same state of the world, reaction is negative for reactor
in that same state of the world, reaction is neutral for reactor
in that same state of the world, reaction is neutral for reactor
SYSTEM COLLAPSES
SYSTEM IS STABLE (BUT NOT COMMUNICATIVE)
COMMUNICATION

Figure 2. Sensory manipulation. As with ritualization, sensory manipulation involves two stages. First, a coercive behaviour emerges. Then, that behaviour may become a signal, and the interaction may become communicative, if the coerced individual (the reactor) gains from being coerced. In contrast, if being coerced is costly, then the system will collapse. If being coerced is neutral, then the system will remain stable, but we cannot label it communicative.

accounts also make it clear that they are not exact mirror images of one another. There are also several small differences, caused by the fact that communication is a dynamic rather than a static game (i.e. one player, the signaller, necessarily acts before the other; in a static game, such as the Prisoner's Dilemma, both players act at the same time). These differences turn out to be important when we ask if one or the other process is likely to be more common in nature.

(c) *Ritualization is likely to be more common than sensory manipulation*

We now compare the exact conditions required for these two processes to occur. For ritualization, the initial conditions required for a cue to emerge are that in some particular state of the world, there is an action for which the actor gains some benefit, independently of any effect it might have upon the reactor; and that there is a reaction for which the reactor gains some benefit. Expressed formally, these conditions are

$$\Pi_{\mathrm{A}}(t_I, r_0) > 0 \qquad (4.1)$$

and

$$\Pi_{\mathrm{R}}(t_I, r_K) > 0. \qquad (4.2)$$

Then, for the cue to become a signal (and the cued behaviour to become a response), we require that the payoff owing to the actor when the reaction is performed is positive:

$$\Pi_{\mathrm{A}}(t_I, r_K) > 0. \qquad (4.3)$$

For sensory manipulation, the initial conditions for coercion to emerge are that there is a particular reaction that, because of some pre-existing mechanism, produces a particular reaction; and that in some particular state of the world, there is a positive payoff to the actor if the reactor performs that same reaction. Expressed formally:

$$q(r_K | a_{\mathcal{J}}) = 1 \qquad (4.4)$$

and

$$\Pi_{\mathrm{A}}(t_I, r_K) > 0. \qquad (4.5)$$

Then, for the coerced behaviour to become a response (and the coercive behaviour to become a signal), we require that the payoff owing to the reactor when they are coerced is positive:

$$\Pi_{\mathrm{R}}(t_I, r_K) > 0. \qquad (4.6)$$

Note that (4.2) is the same as (4.6), and that (4.3) is the same as (4.5). In other words, both processes require that both participants benefit (this follows from the way in which communication is defined; see §2). The difference lies in which of these conditions is necessary for the first stage, in which cueing or coercion emerges; and which is necessary for the second stage, in which the cue/coercion becomes a signal. Each process also has an additional condition that is necessary to trigger the first stage. However, in the case of ritualization, the additional condition, (4.1), is already entailed by (4.3), so all that is required are conditions (4.2) and (4.3). However, for sensory manipulation, the additional condition, (4.4), really is an additional condition.

In other words, in ritualization, the condition necessary for a cue to become a signal is already partially satisfied by the condition necessary for a cue to emerge in the first place. This is not, however, true of coercion: condition (4.4) has no bearing on condition (4.6). So with ritualization, the (proto-)signal is likely to be 'honest' by virtue of the way in which it emerges: if it did not accurately reflect some aspect of the world that is pertinent to the reactor (e.g. from the example used above, urine reveals the presence of an animal), then the reactor would not have evolved to attend to it in the first place. However, there is no similar requirement for sensory manipulation. Here, actors evolve to manipulate reactors, but this manipulation may not be honest—and if it is not, then reactors will evolve to ignore actors, and the system will collapse. This difference between the two processes may explain why, in the empirical literature, ritualization is observed to be more common than sensory manipulation.

The point is not that cues will always become signals. For example, if the urine revealed the location of the animal to potential predators, then once it became a cue the animal would be under a selection pressure not to urinate so conspicuously. The point is instead that the likelihood that cues will become signals is greater than the likelihood that coerced behaviours become responses. This is because in ritualization the actor already receives some benefit from the action, independent of the effects the action has on the reactor—but with sensory manipulation there is no equivalent foundation: the reactor does not receive any prior benefit.

5. IS THERE A THIRD ROUTE TO COMMUNICATION?

Thus far, we have discussed only two possible answers to the question 'how do communication systems emerge?'. In this section, we ask whether these two answers are exhaustive, or whether there is an additional, third route to communication. In particular, we ask whether it is possible to go from a state of non-communication to a state of communication *without* first passing through either a state of cueing or a state of coercion.

Since signal and response are interdependent, this direct emergence would require that signal and response come into existence simultaneously. With natural selection this is possible but unlikely, since it requires simultaneous, complementary mutations in actor and reactor. However, we can imagine how it might occur in other domains. Specifically: if the mechanisms that determine behaviour are such that changes in the actor's mechanism trigger immediate and complementary changes in the reactor's mechanism, it would be possible for communication to emerge directly from non-communication: chicken and egg at the same time. More precisely, reactors must have mechanisms that allow them to recognize that a novel behaviour is designed to be a signal; and signallers must have mechanisms that allow them to create signals that have the features that allow receivers to recognize them as such (figure 3).

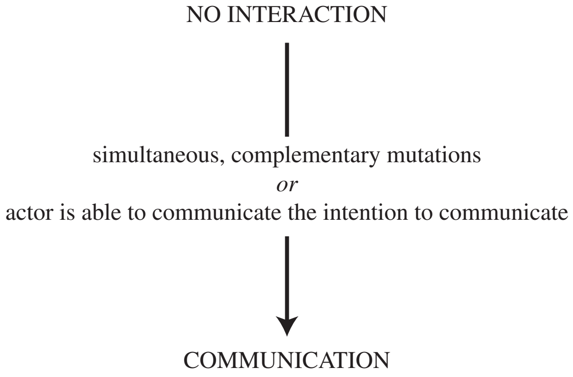
NO INTERACTION
simultaneous, complementary mutations
or
actor is able to communicate the intention to communicate
COMMUNICATION

Figure 3. The direct route to communication. It is possible for communication to emerge directly from a state of no interaction, without going via cues or coercion, but only if the individuals involved possess mechanisms that allow them to recognize the functionality of each other's behaviour. In short, (proto-)signallers must communicate that they have an intention to communicate. This ability may be limited to humans.

This is quite a specific requirement, but humans possess such a mechanism, in their capacity to attribute intentions to others' behaviour [15]. In short, humans are able to make it manifest to their audience that they wish to communicate with them [16–18]. For example, if I wish to request more wine, I can do this simply by tilting my empty glass towards my host in a particular way. Not only does this inform my host that I wish for more wine, it also informs her that it is my intention to inform her that I wish for more wine. As such, the behaviour is a signal (about both my desire for wine, and about my intention to communicate the same), yet, crucially, there is no preceding cue or coercive behaviour here. Instead, communication emerges directly and immediately. This is, then, a direct route to communication, different to both ritualization and sensory manipulation.

The same point is illustrated by recent empirical work in which pairs of interacting human participants play simple computer games that involve coordinating their behaviour with one another, and doing so successfully requires that they find a way to reveal when their behaviour is intended to be communicative [19,20]. These challenges prove difficult, but participants are able to overcome them, and in doing so they demonstrate that humans are able to establish a communication system directly, without going via cues and coercive behaviours [19]. It is possible that other species, in particular some non-human primates, possess the cognitive mechanisms that allow this third route to emerge, but there is currently no demonstration of this. Indeed, whether this is the case is a key empirical question.

Another way to interpret this claim is to ask how it relates to the result derived in §3*b*: that a state of non-interaction is evolutionarily stable. That result showed that communication cannot emerge if changes in strategy must be unilateral. However, if both participants are able to change their behaviour simultaneously, and in complementary ways, then a state of not communicating is no longer stable, and communication can emerge directly. Such simultaneous changes are likely to be rare in natural selection, but in other domains they are possible. In particular, this is what happens with human dyads, because as signallers humans are able to construct their behaviour such that it reveals the intentions behind it, and as receivers they are able to recognize those same intentions.

6. DISCUSSION

Our model illustrates three main points with respect to the emergence of communication. First, a state of non-interaction is evolutionarily stable—and so we cannot simply assume that communication will evolve, even if it would be in both parties' interests. Second, of the two ways in which animal signals are known to evolve, ritualization (cue first) is likely to be the more common, because the prerequisites for it to occur are less restrictive than they are for sensory manipulation (coercion first). This is because cues are already 'honest', in that they reliably reflect some aspect of the world; but coercive behaviours are not. Third, humans (and perhaps only humans) are able to develop communication systems in a more direct way, by virtue of their ability to make their communicative intentions manifest to others.

Our results also demonstrate the utility of a functional approach to communication. There is currently an ongoing interdisciplinary discussion about how best to conceptualize communication, and about how we might develop a consilient account of communication [21]. Some participants in this discussion have promoted information-theoretic approaches [22]. Others have argued that functionality and influence provide the foundations of communication [4,23]. It may turn out that functional perspectives and information-centric perspectives are compatible with one another [24]—but at the same time, our results demonstrate that a functional approach yields real insights into several aspects of how communication systems emerge. It is not clear if the same insights can be derived from other perspectives.

Nothing in our framework is specific to the process of natural selection, and our results hence apply more widely. For example, the framework could also be used to describe ontogenetic ritualization, in which pairs of interacting primates develop communicative conventions that are unique to that dyad [6,7]. Indeed, it is only because of the general nature of our terms and definitions that we have been able to compare the emergence of communication in human dyads with the evolution of animal signals, and hence make the claim (in §5) that the human ability to reveal and detect communicative intentions provides a third route to communication.

Finally, we wish to reiterate the point with which we began: that communication is not a trait possessed by an individual, but rather the consequence of a certain type of interaction; specifically, one that has interdependent functionality. It is only because we adopted and built upon a definition of communication that captured this fact that we have been able to derive the results that we have. We believe this approach captures the essence of communication, and is hence both fruitful and accurate.

T.C.S.-P. acknowledges financial support from the Leverhulme Trust; R.A.B. from Research Councils UK; A.G. from Balliol College and the Royal Society; and S.A.W. from the European Research Council.

REFERENCES

1 Krebs, J. R. & Dawkins, R. 1984 Animal signals: mind-reading and manipulation. In *Behavioural ecology: an evolutionary approach* (eds J. R. Krebs & N. B. Davies), pp. 380–402, 2nd edn. Oxford, UK: Blackwell.

2 di Paolo, E. A. 1997 An investigation into the evolution of communication. *Adapt. Behav.* **6**, 285–324. (doi:10.1177/105971239700600204)

3 Maynard Smith, J. & Harper, D. G. C. 2003 *Animal signals*. Oxford, UK: Oxford University Press.

4 Scott-Phillips, T. C. 2008 Defining biological communication. *J. Evol. Biol.* **21**, 387–395. (doi:10.1111/j.1420-9101.2007.01497.x)

5 Bradbury, J. W. & Vehrencamp, S. L. 2011 *Principles of animal communication*. Sunderland, MA: Sinauer Associates.

6 Tomasello, M. & Call, J. 1997 *Primate cognition*. Oxford, UK: Oxford University Press.

7 Tomasello, M. & Zuberbühler, K. 2002 Primate vocal and gestural communication. In *The cognitive animal* (eds M. Bekoff, C. Allen & M. Burghardt), pp. 293–299. Cambridge, MA: MIT Press.

8 Lotem, A., Wagner, R. H. & Balshine-Earn, S. 1999 The overlooked signalling component of nonsignaling behavior. *Behav. Ecol.* **10**, 209–212. (doi:10.1093/beheco/10.2.209)

9 Donaldson, M. C., Lachmann, M. & Bergstrom, C. T. 2007 The evolution of functionally referential meaning in a structured world. *J. Theor. Biol.* **246**, 225–233. (doi:10.1016/j.jtbi.2006.12.031)

10 Godfray, H. C. J. 1991 Signalling of need by offspring to their parents. *Nature* **352**, 328–330. (doi:10.1038/352328a0)

11 Godfray, H. C. J. & Johnstone, R. A. 2000 Begging and bleating: the evolution of parent–offspring signalling. *Phil. Trans. R. Soc. Lond. B* **355**, 1581–1591. (doi:10.1098/rstb.2000.0719)

12 Lorenz, K. 1970 *Studies in animal and human behaviour*, vol. 1. London, UK: Methuen.

13 Vahed, K. 1998 The function of nuptial feeding in insects—review of empirical studies. *Biol. Rev.* **73**, 43–78. (doi:10.1017/S0006323197005112)

14 Thornhill, R. 1976 Sexual selection and nuptial feeding in *Bittacus apicalis* (Insecta: Mecoptera). *Am. Nat.* **110**, 529–548. (doi:10.1086/283089)

15 Csibra, G. & Gergely, G. 2007 'Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. *Acta Psychol.* **124**, 60–78. (doi:10.1016/j.actpsy.2006.09.007)

16 Grice, H. P. 1975 Logic and conversation. In *Syntax and semantics III: Speech acts* (eds P. Cole & J. Morgan), pp. 41–58. New York, NY: Academic Press.

17 Sperber, D. 2000 Metarepresentations in an evolutionary perspective. In *Metarepresentations: an interdisciplinary perspective* (ed. D. Sperber), pp. 117–137. Oxford, UK: Oxford University Press.

18 Tomasello, M. 2008 *Origins of human communication*. Cambridge, MA: MIT Press.

19 Scott-Phillips, T. C., Kirby, S. & Ritchie, G. R. S. 2009 Signalling signalhood and the emergence of communication. *Cognition* **113**, 226–233. (doi:10.1016/j.cognition.2009.08.009)

20 de Ruiter, J. P., Noordzij, M. L., Newman-Norland, S., Newman-Norland, R., Hagoort, P., Levinson, S. C. & Toni, I. 2010 Exploring the cognitive infrastruture of communication. *Interact. Stud.* **11**, 51–77. (doi:10.1075/is.11.1.05rui)

21 Stegmann, U. E. (ed.) Forthcoming. *Animal communication theory: information and influence*. Cambridge, UK: Cambridge University Press.

22 Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K. & Hammerschmidt, K. 2010 The central importance of information in studies of animal communication. *Anim. Behav.* **80**, 3–8. (doi:10.1016/j.anbehav.2010.04.012)

23 Rendall, D., Owren, M. J. & Ryan, M. J. 2009 What do animal signals mean? *Anim. Behav.* **78**, 233–240. (doi:10.1016/j.anbehav.2009.06.007)

24 Carazo, P. & Font, E. 2010 Putting information back into biological communication. *J. Evol. Biol.* **23**, 661–669. (doi:10.1111/j.1420-9101.2010.01944.x)

Supplementary information: *The nature of communication.*

Because it emphasises interdependence, our definition of communication (stated in the main paper) may appear to suggest that communication is an inherently cooperative act. Yet many instances of communication are antagonistic: two dogs that bare their teeth in an aggressive contest over, say, food, are communicating with one another about their relative fighting abilities, but are otherwise engaged in a hostile interaction. This apparent problem makes clear the need to distinguish between different types of cooperation involved in communication. Here, we adopt a three-way distinction [26; see table S1]. This will also allow us to specify the goals of our study more precisely.

[table S1 about here]

First, signals and responses must be calibrated to one another: signaller and receiver must agree upon what a signal 'means'. (For example, in human language, speaker and listener must agree that "dog" refers to canine animals, and not to feline animals; otherwise meaningful communication cannot even take place.) We call this *communicative cooperation*. This type of cooperation is necessary for communication to take place in the first place; without it, the interaction is not communicative.

Second, signals may or may not be honest; they may or may not reliably correlate with some feature of the world. We call this *informative cooperation*. If signals are not informatively cooperative, then the system is likely to collapse. How this outcome is avoided, and hence how communication systems can remain evolutionarily stable, is the defining problem of animal signalling theory [3, 27-29]. Note that informative and communicative cooperation are dependent on one another. A system that is communicatively uncooperative cannot be honest (or even dishonest), since the signals do not yet mean anything. Similarly, a system that is informatively uncooperative will soon collapse, and hence there will be no communication to be cooperative about. Consequently, both informative and communicative cooperation are necessary for communication to be evolutionarily stable.

Finally, communication may occur in cooperative or competitive contexts (e.g. building a nest together vs. fighting over territory). We call this *material cooperation*. Material cooperation is not necessary for communication to exist, nor for it to be stable. Communication can be materially uncooperative, while at the same time being communicatively and informatively cooperative, for example between teeth-baring dogs.

These distinctions allow us to state the goals of the present study more specifically: we are investigating the *origins of communicative cooperation*. There are many models and empirical studies of informative cooperation [see 3, 29 for reviews]. There are also many models, both mathematical and computational, that investigate how pre-existing signal forms become attached to particular meanings; that is, how a communication system might move from a state of communicative non-cooperation to a state of communicative cooperation [e.g. 30-33]. Still other theoretical work has shown that the evolution of communication may depend upon what behavioural strategy is pursued by the participants prior to communication [34]. However previous research has not systematically investigated the *origins* of communicative cooperation i.e. how the necessary interdependence between signals and responses might emerge in the first place.

type of cooperation	*gloss*	*necessary for communication to be stable?*
communicative	Are signals and responses calibrated to one another? (Do signaller and receiver agree on what a signal 'means'?)	Yes
informative	Are signals honest? (Do they reliably correlate with some feature of the world?)	Yes
material	Is communication used in cooperative or competitive contexts (e.g. building a nest together vs. fighting over territory)?	No

Table S1: *The different types of cooperation involved in communication.*
Because our framework emphasises the interdependence of signals and responses, it stresses the inherently cooperative nature of communication. Yet many communicative scenarios are antagonistic – and so it is important to distinguish between three different types of cooperation that are involved in communication. Only the first two in this table (communicative and

informative cooperation) are necessary for evolutionary stability. For further discussion see [26].

References

3. Maynard Smith, J., & Harper, D. G. C. (2003). *Animal Signals*. Oxford: Oxford University Press.

26. Scott-Phillips, T. C. (2010). Animal communication: Insights from linguistic pragmatics. *Animal Behaviour, 79*(1), e1-e4.

27. Grafen, A. (1990). Biological signals as handicaps. *Journal of theoretical biology, 144*, 517-546.

28. Grafen, A. (1990). Sexual selection unhandicapped by the Fisher process. *Journal of Theoretical Biology, 144*, 473-516.

29. Searcy, W. A., & Nowicki, S. (2007). *The Evolution of Animal Communication*. Princeton, NJ: Princeton University Press.

30. Hurford, J. R. (1989). Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua, 77*, 187-222.

31. Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences, 96*, 8028-8033.

32. Nowak, M. A., Plotkin, J. B., & Jansen, V. A. (2000). The evolution of syntactic communication. *Nature, 404*, 495-498.

33. Skyrms, B. (2010). *Signals*. Oxford: Oxford University Press.

34. Bradbury, J. W., & Vehrencamp, S. L. (2000). Economic models of animal communication. *Animal Behaviour, 59*, 259-268.

Supplementary Information: *Proof that* ($\boldsymbol{P}_{null}$, $\boldsymbol{Q}_{null}$) *is evolutionarily stable.*

We show that any unilateral change in strategy will not increase the payoff of either player. This shows that ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$) is a Nash equilibrium. It then follows that not communicating is evolutionarily stable, since all Nash equilibria in role-asymmetric games are evolutionarily stable strategies [35]. (Role-asymmetric games are those where the players have different roles, which is the case here.)

For ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$),

$$w_A(P_{null}, Q_{null}) = \sum_{t\in T} \varphi(t)\Pi_A(t, r_0)$$

What happens if the actor changes strategy to $\mathbf{P}'$, where $\mathbf{P}' = \mathbf{P}$, except that $p(a_J \mid t_I) \neq 0$ for some specific K, $J \neq 0$, and $p(a_0 \mid t_I) = 1 - p(a_J \mid t_I)$? Now we have

$$w_A(P', Q_{null}) = \sum_{t\in T} \varphi(t)\Pi_A(t, r_0) - \varepsilon(a_J)\varphi(t_I)p(a_J \mid t_I)$$

Since $\varepsilon(a_J)$, $\varphi(t_I)$ and $p(a_J \mid t_I)$ are all positive, then $w_A(\boldsymbol{P}_{null}, \boldsymbol{Q}_{null}) > w_A(\boldsymbol{P}', \boldsymbol{Q}_{null})$, and therefore $\mathbf{P}'$ is a strictly worse strategy for the actor than $\mathbf{P}_{\text{null}}$.

Similarly, at ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$),

$$w_R(P_{null}, Q_{null}) = \sum_{t\in T} \varphi(t)\Pi_R(t, r_0)$$

What happens if the reactor changes strategy to $\mathbf{Q}'$, where $\mathbf{Q}' = \mathbf{Q}$, except that $q(r_K \mid a_J) \neq 0$ for some specific K, $J \neq 0$, and $q(r_0 \mid a_J) = 1 - q(r_K \mid a_J)$? Then $w_R(\boldsymbol{P}_{null}, \boldsymbol{Q}') = w_R(\boldsymbol{P}_{null}, \boldsymbol{Q}_{null})$, since actor always performs a_0, and never a_J, so r_K never actually occurs.

Therefore neither player has an incentive to unilaterally change strategy at ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$), and hence ($\mathbf{P}_{\text{null}}$, $\mathbf{Q}_{\text{null}}$) is a Nash equilibrium.

Reference

35. Selten, R., 1980, A note on evolutionarily stable strategies in asymmatric animal conflicts. *Journal of Theoretical Biology, 84*, 93-101.